Available online at www.sciencedirect.com
SCIENCE DIRECT®
Bioorganic & Medicinal Chemistry 13 (2005) 2109–2120
Bioorganic & Medicinal Chemistry

A novel method of estimation of lipophilicity using distance-based topological indices: dominating role of equalized electronegativity

Vijay K. Agrawal,[a] Madhu Gupta,[b] Jyoti Singh[a] and Padmakar V. Khadikar[c,*]

[a]*QSAR and Computer Chemical Laboratories, A.P.S. University, Rewa 486 003, India*
[b]*Department of Chemistry, MMH P.G. College, Ghaziabad 201 001, India*
[c]*Research Division, Laxmi Fumigation and Pest Control Pvt. Ltd, 3 Khatipura, Indore 452 007, India*

Received 18 December 2004; revised 1 January 2005; accepted 4 January 2005

Abstract—Attempt is made to propose yet another method of estimating lipophilicity of a heterogeneous set of 223 compounds. The method is based on the use of equalized electronegativity along with topological indices. It was observed that excellent results are obtained in multiparametric regression upon introduction of indicator parameters. The results are discussed critically on the basis various statistical parameters.

1. Introduction

The partitioning of organic compounds acting as drugs between aqueous and lipophilic phase is of utmost importance for drug potency. Lipophilicity is an important endpoint used extensively in medicinal chemistry, drug design, pharmacy, and environmental toxicity in predicting biological and hazardous effects of chemicals.[1] Consequently, the development of new methods for the quantitative estimation of the lipophilicity of organic compounds is extremely important. It is well known and well established that the useful measure of lipophilicity/hydrophobicity of organic compounds acting as drugs is then passion coefficient between *n*-octanol and water.[1] No other physicochemical property has attracted as much interest in quantitative structure–activity relation (QSAR) studies as lipophilicity (synonymously called hydrophobicity, any difference between both these terms is only a semitic nemicity; the opposite of lipophilicity is hydrophobicity) generally expressed as $\log P$.

Due to predominant role of lipophilicity ($\log P$) in biological activity and despite the large number of papers already published,[2–5] lipophilicity still constitute a challenging subject for further investigation. Although a great deal is known about the experimental determination and the calculation of $\log P$, the most widely accepted measure of lipophilicity, the problem of estimating the $\log P$ values of new compounds is far from being solved in many cases.

Attempts to relate the chemical or biological properties of organic compounds with their structure, a easily measured molecular descriptors has led to a proliferation of methods and indices.[2–10] At a fundamental level, indices include pure graph theoretical invariant commonly called topological indices; such as Wiener (W),[11] Szeged (Sz),[12,13] Randic ($^{m}\chi$),[14] Kier and Hall ($^{m}\chi^{V}$)[15,16] indices, etc. There have been a large number of other less fundamental indices, which have attempted to capture the essence of molecular shape, reactivity, and polarity. Such indices are successfully used in modeling physicochemical properties and physiological activity of organic compounds acting as drugs. However, when more complex physicochemical or biological properties such as lipophilicity/hydrophobicity or drug effectiveness are to be modeled then QSAR/QSPR indices need to reflect not only molecular bulk but also the bonding interaction between atoms in molecules and of

Keywords: Lipophilicity; Topological index; Equalized electronegativity; Regression analysis; QSAR.
* Corresponding author. Tel.: +91 731 2531906; fax: +91 731 2763618; e-mail addresses: vkagrawal@lycos.com; pvkhadikar@rediffmail.com

0968-0896/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmc.2005.01.003

molecules with the external environment. Gombar and Enslein[17] used Kier and Hall electro-topological indices[18] for this purpose and obtained excellent agreement between calculated and observed lop. However, their method had a 23% calling of data. Also, that their data set modeled with 363 independent variables. Similarly, very recently we have proposed novel estimation of $\log P$ using newly introduced PI (Padmakar–Ivan) index.[19,20] Also, we have modeled lipophilicity using information theoretic indices[21] as well as some of the distance-based topological indices. The primary aim of our earlier investigation was for providing topological (structural) evidences for modeling lipophilicity ($\log P$) and not to provide new method for the calculation of $\log P$.

Our earlier attempts of modeling lipophilicity were concerned with a limited set of compounds. The wider use of our methodology is only possible when it successfully applied to a large set of heterogeneous compounds. This was, therefore, the primary aim of the present investigation in that we have chosen a large set of 223 heterogeneous compounds and modeled their $\log P$ using a larger set of distance-based topological indices. Along with the topological indices, we have also chosen equalized electronegativity (χ_{eq})[22] for discussing the effect on lipophilicity ($\log P$) due to electronegativity. The results as discussed below show that excellent model is obtained when these indices are coupled with some indicator parameters.

2. Result and discussion

The set of 223 compounds used in the present investigation is given in Table 1. Table 1 also contains $\log P$ values and assumed indicator parameters. The calculated values of topological indices : W, Sz, $^1\chi$, J, $\log$RB as well as of χ_{eq} are recorded in Table 2. The correlations between the topological indices and their correlations with $\log P$ are presented in Table 3.

A preliminary regression analysis[23] has indicated that for modeling $\log P$ excellently the data set of 223 compounds is to be split into two sub-groups: one containing 210 compounds and the other containing 13 compounds. This is perhaps due to the fact that the compounds belonging to these sub-groups have different types of mechanism or this may be due to the regression procedure adopted by us. In fact, due to their occurrence as outliners in the most significant regression expression proposed the second subset of 13 compounds is generated. Whatever may be the reason deletion of 13 compounds, thus forming a sub-set, represents only 5% calling of data.

Before discussing the results and proposing statistically significant models it is necessary to mention that the basic which determined the topological conditioning of the reactivity is the principle of '*molecular structure*', according to which, molecules are considered as isolated objects, possessing a relatively rigid and permanent location of nuclei (atoms), joined each other by electronic forces (chemical bonds), which are highly specific and strongly localized. Hence molecules are assumed to have a structure, which conditions their physiochemical properties and physiological activities. Therefore, as a consequence of this principle, it is hardly surprising that the most of '*natural*' topological indices (distance-based topological indices) we used herein renders much satisfactory results. It is worth mentioning that although current topological indices can be compared in a nearly simple way, they are chosen on the grounds that in spite of their high co-linearity (Table 4) they have different information contents.

In modeling $\log P$ of the heterogeneous set of 223 compounds and to arrive at most significant model, we have chosen stepwise regression analysis and used maximum R^2-method. The results obtained as discussed below.

Using simple as well as bi-parametric regressions, we could not obtain any statistically significant regression expression for modeling $\log P$ of the heterogeneous set of 223 compounds. We have, therefore, carried out tri-parametric regression analyses. Among the several tri-parametric regressions attempted the one containing $^1\chi$, Ip_2 and Ip_3 was found the best:

$$\begin{aligned}\log P &= 0.4467 + 0.3672(\pm 0.0594)^1\chi \\ &\quad + 1.4262(\pm 0.2380)\mathrm{Ip}_2 - 1.3032(\pm 0.2354)\mathrm{Ip}_3 \\ n &= 223, \quad \mathrm{Se} = 1.2187, \quad R = 0.7740, \\ R^2_{\mathrm{A}} &= 0.5935, \quad F = 109.053, \quad Q = 0.6351\end{aligned} \tag{1}$$

Here, and thereafter, n is the number of compounds, Se is the standard error of estimation, R is multiple correlation coefficient, R^2_{A} is the adjusted R squared, and F is the F-statistics, Q is the quality factor, which is the ratio of R/Se.

In the above Eq. 1 the coefficient of $^1\chi$ and Ip_2 terms are positive while that of Ip_3 is negative. This means that number of atoms, first-order branching and presence of aromatic nucleus are favorable for the exhibition of $\log P$. The presence of –OH group on the other hand is not favorable for this purpose.

When Sz index is added to the above Eq. 1 we obtained the tetra-parametric regression expression having slightly better statistics. The corresponding expression using W in place of Sz was less significant. The tetra-parametric regression expression contains $^1\chi$, Sz, Ip_2 and Ip_3 is found as below:

$$\begin{aligned}\log P &= 0.8088 + 0.2159(\pm 0.1081)^1\chi \\ &\quad + 9.2761 \times 10^{-4}(\pm 5.5439 \times 10^{-4})\mathrm{Sz} \\ &\quad + 1.5365(\pm 0.2460)\mathrm{Ip}_2 - 1.2785(\pm 0.2349)\mathrm{Ip}_3 \\ n &= 223, \quad \mathrm{Se} = 1.2138, \quad R = 0.7772, \\ R^2_{\mathrm{A}} &= 0.5968, \quad F = 83.162, \quad Q = 0.6403\end{aligned} \tag{2}$$

Table 1. The compounds used, their lipophilicity (log *P*) and indicator parameters

S.N.	Compounds name	log *P*	Ip_1	Ip_2	Ip_3
1.	Chlorodifluoromethane	1.08	1	0	0
2.	Dichlorodifluoromethane	2.16	1	0	0
3.	Fluorotrichloromethane	2.53	1	0	0
4.	Carbon tetrachloride	2.83	1	0	0
5.	Dichloromethane	1.25	1	0	0
6.	Formaldehyde	0.35	0	0	0
7.	Formamide	−1.51	0	0	0
8.	Methanol	−0.77	0	0	1
9.	Tetrachloroethylene	3.40	1	0	0
10.	Hexafluoroethane	2.00	1	0	0
11.	Trichloroethylene	2.61	1	0	0
12.	Pentachloroethane	3.22	1	0	0
13.	Acetylene	0.37	0	0	0
14.	*trans*-1,2-Dichloroethylene	2.09	1	0	0
15.	Dichloroacetic acid	0.92	1	0	0
16.	1,1,2,2-Tetrachloethane	2.39	1	0	0
17.	1,1-Difluoroethylene	1.24	1	0	0
18.	Bromoacetic acid	0.41	1	0	0
19.	1,1,1-Trichloroethane	2.49	1	0	0
20.	Acetonitrile	−0.34	0	0	0
21.	Ethene	1.13	0	0	0
22.	1,2-Dibromoethane	1.96	1	0	0
23.	1,2-Dichloroethane	1.48	1	0	0
24.	Bromoethane	1.61	1	0	0
25.	Ethylchloride	1.43	1	0	0
26.	2-Fluoroethnol	−0.76	1	0	1
27.	Glycine	−3.21	0	0	0
28.	Ethane	1.81	0	0	0
29.	Ethanol	−0.31	0	0	1
30.	Ethylene glycol	−1.36	0	0	1
31.	Ethanolamine	−1.31	0	0	1
32.	Ethylene diamine	−2.04	0	0	0
33.	Hexafluoro-2-propanol	1.66	1	0	1
34.	Acrylonitrile	0.25	0	0	0
35.	Tetrahydrofuran	0.46	0	0	0
36.	Ethyl acetate	0.73	0	0	0
37.	Butyric acid	0.79	0	0	0
38.	1,4-Dioxane	−0.27	0	0	0
39.	1-Chlorobutane	2.64	1	0	0
40.	4-Chlorobutanol	0.85	1	0	1
41.	Dimethylacetamide	−0.77	0	0	0
42.	*n*-Butane	2.76	0	0	0
43.	Butanol	0.88	0	0	1
44.	Diethyl ether	0.89	0	0	0
45.	1,3-Butanediol	−0.83	0	0	1
46.	2,3-Butanediol	−0.92	0	0	1
47.	Diethylamine	0.58	0	0	0
48.	1-Butylamine	0.73	0	0	0
49.	Diethanolamine	−1.43	0	0	1
50.	2-Cyanopyrimidine	−0.31	0	0	0
51.	2-Bromopyridine	1.42	1	0	0
52.	1,4-Pentadiene	2.48	0	0	0
53.	Cyclopentane	3.00	0	0	0
54.	Diethyl ketone	0.99	0	0	0
55.	Ethyl propionate	1.24	0	0	0
56.	Diethyl carbonate	1.21	0	0	0
57.	1-Chloropentane	2.52	1	0	0
58.	Neopentane	3.11	0	0	0
59.	*n*-Pentane	3.39	0	0	0
60.	2-Pentanol	1.19	0	0	1
61.	Oxazole	0.12	0	0	0
62.	Propane	0.94	0	0	0
63.	Imidazole	−0.08	0	0	0
64.	Acrolein	−0.38	0	0	0
65.	Acrylic acid	0.35	0	0	0
66.	Cyclopropane	1.72	0	0	0
67.	Propene	1.77	0	0	0
68.	1-Bromo-3-chloropropane	2.18	1	0	0
69.	Acetone	−0.24	0	0	0
70.	Allyl alcohol	0.17	0	0	1
71.	Propionic acid	0.33	0	0	0
72.	3-Chloro-1-propanol	0.50	1	0	1
73.	3-Fluoropropanol	−0.28	1	0	1
74.	Diethylformamide	−1.01	0	0	0
75.	Propane	2.36	0	0	0
76.	2-Propanol	0.05	0	0	1
77.	Propanol	0.25	0	0	1
78.	1,2-Propanediol	−0.92	0	0	1
79.	2,6-Dichloropyrazine	1.53	1	0	0
80.	2-Fluoropyrimidine	0.02	1	0	0
81.	Pyrazine	−0.23	0	0	0
82.	Uracil	−1.07	0	0	0
83.	Furan	1.34	0	0	0
84.	Pyrrole	0.75	0	0	0
85.	Cytosine	−1.73	0	0	0
86.	1,3-Butadiene	1.99	0	0	0
87.	2-Butyne	1.46	0	0	0
88.	4-Bromo-1-butene	2.53	1	0	0
89.	4,4,4-Trifluorobutanol	0.71	1	0	1
90.	Butyronitrile	0.53	0	0	0
91.	1-Butene	2.40	0	0	0
92.	2-Butene	2.33	0	0	0
93.	Methylpropene	2.34	0	0	0
94.	2-Butanone	0.29	0	0	0
95.	Butanal	0.88	0	0	0
96.	Bromobenzene	2.99	1	1	0
97.	*p*-Bromophenol	2.59	1	1	1
98.	Chlorobenzene	2.89	1	1	0
99.	Fluorobenzene	2.27	1	1	0
100.	*p*-Nitrophenol	1.91	0	1	1
101.	Benzene	2.13	0	1	0
102.	1,3-Cyclohexadiene	2.47	0	0	0
103.	1,4-Ccyclohexadiene	2.30	0	0	0
104.	1-Hexyn-5-one	0.58	0	0	0
105.	1,5-Hexadiene	2.87	0	0	0
106.	1-Hexyne	2.73	0	0	0
107.	Cyclohexene	2.86	0	0	0
108.	1-Hexene	3.39	0	0	0
109.	Cyclohexane	3.44	0	0	0
110.	Methyl-cyclopentane	3.37	0	0	0
111.	Cyclohexanol	1.23	0	0	1
112.	*n*-Butyl acetate	1.78	0	0	0
113.	Cyclohexyl-amine	2.74	0	0	0
114.	*N*,*N*-Diethylacetamide	0.34	0	0	0
115.	*n*-Hexane	3.90	0	0	0
116.	2-Hexanol	1.76	0	0	1
117.	3,3-Dimethyl-1-butanol	1.48	0	0	1
118.	2-Decanone	1.65	0	0	0
119.	3-Hexanol	1.65	0	0	1
120.	Diisopropylether	1.52	0	0	0
121.	Butoxyethanol	0.83	0	0	1
122.	Hexachlorobenzene	5.73	1	1	0
123.	Pentachlorobenzene	5.18	1	1	0
124.	1,2,3,4-Tetrachlorobenzene	4.64	1	1	0
125.	1,2,3,5-Tetrachlorobenzene	4.66	1	1	0
126.	1,2,4,5-Tetrachlorobenzene	4.60	1	1	0
127.	1,2,4-Trichlorobenzene	4.02	1	1	0
128.	1,3,5-Trichlorobenzene	4.19	1	1	0
129.	1,2-Dibromobenzene	3.64	1	1	0
130.	1,4,Dibromobenzene	3.79	1	1	0
131.	1,3-Dichlorobenzene	3.53	1	1	0

(*continued on next page*)

Table 1 (*continued*)

S.N.	Compounds name	log *P*	Ip_1	Ip_2	Ip_3
132.	1,2-Dichlorobenzene	3.43	1	1	0
133.	1,4-Dichlorobenzene	3.44	1	1	0
134.	2-Nitro-5-fluorophenol	1.91	1	1	1
135.	3-Acetylpyridine	0.34	0	0	0
136.	4-Acetylpyridine	1.31	0	0	0
137.	Toluene	2.73	0	1	0
138.	Anisole	2.11	0	1	0
139.	Benzyl alcohol	1.10	0	1	1
140.	Cresol	1.96	0	1	1
141.	Cycloheptane	4.00	0	0	0
142.	1-Methylcyclohexanol	1.33	0	0	1
143.	*n*-Heptane	4.66	0	0	0
144.	Heptyl-amine	2.57	0	0	0
145.	Benzofuran	2.67	0	1	0
146.	Acetophenone	1.58	0	1	0
147.	Ethylbenzene	3.15	0	1	0
148.	*m*-Xylene	3.20	0	1	0
149.	*o*-Xylene	3.12	0	1	0
150.	*p*-Xylene	3.15	0	1	0
151.	Caffeine	−0.07	0	0	0
152.	1,2-Dimethoxybenzene	1.60	0	1	0
153.	*N*,*N*-Dimethylaniline	2.31	0	1	0
154.	*n*-Octane	5.18	0	0	0
155.	Diisobutyl ether	2.78	0	0	0
156.	Diethylene glycol butyl ether	0.56	0	0	1
157.	1,2,4-Trimethylbenzene	3.78	0	1	0
158.	Allylbenzene	3.23	0	1	0
159.	Etylbenzoate	2.64	0	1	0
160.	1,2,3-Trimethylbenzene	3.66	0	1	0
161.	1,3,5-Trimethylbenzene	3.42	0	1	0
162.	Isopropylbenzene	3.66	0	1	0
163.	*n*-Propylbenzene	3.72	0	1	0
164.	*N*,*N*-Dimethyltoluamine	2.85	0	1	0
165.	2-(5-NO_2-2-Furfurilidine)-pyrimidine	1.34	0	0	0
166.	Naphthalene	3.30	0	1	0
167.	Dimethylteraphthalate	2.25	0	1	0
168.	*p*-Propyl-benzoic acid	3.42	0	1	0
169.	4-*tert*-Butyl nitrobenzene	3.89	0	1	0
170.	1,2,4,5-Tetramethylbenzene	4.00	0	1	0
171.	*p*-Diethylbenzene	4.45	0	1	0
172.	*tert*-Butylbenzene	4.11	0	1	0
173.	*n*-Butylbenzene	4.38	0	1	0
174.	Diethylacetal	0.84	0	0	0
175.	Ethylene glycol isobutyl ether	0.75	0	0	1
176.	Dipropylene glycol	−0.67	0	0	1
177.	Dimethylbutylamine	1.76	0	0	0
178.	Dipropyl-amine	1.67	0	0	0
179.	Trio-ethyl-amine	1.45	0	0	0
180.	3,5-Dinitrobenzoic acid	1.55	0	1	0
181.	Benzonitrile	1.56	0	1	0
182.	Benzimidazole	1.32	0	1	0
183.	Benzaldehyde	1.48	0	1	0
184.	Benzoic acid	1.87	0	1	0
185.	Benzyl chloride	2.30	1	1	0
186.	2-Acetylpyridine	0.85	0	0	0
187.	Pentamethylbenzene	4.56	0	1	0
188.	2,2,4,5-Tetrachlorobiphenyl	5.78	1	1	0
189.	2,3,4,5-Tetrachlorobiphenyl	6.41	1	1	0
190.	2,2,4-Trichlorobiphenyl	5.76	1	1	0
191.	2,4,5-Trichlorobiphenyl	5.90	1	1	0
192.	2,4,6-Trichlorobiphenyl	5.71	1	1	0
193.	2,5-Dichlorobiphenyl	5.16	1	1	0
194.	2,6-Dichlorobiphenyl	4.98	1	1	0
195.	2-Chlorobiphenyl	4.53	1	1	0
196.	Acenaphthene	3.92	0	1	0
197.	Biphenyl	4.01	0	1	0
198.	Benzyl ether	4.21	0	1	0
199.	1,3-Dimethylnaphthalene	4.42	0	1	0
200.	1,4-Dimethylnaphthalene	4.37	0	1	0
201.	1,5-Dimethylnaphthalene	4.38	0	1	0
202.	2,3-Dimethylnaphthalene	4.40	0	1	0
203.	2,6-Dimethylnaphthalene	4.31	0	1	0
204.	Diethylphthalate	2.47	0	1	0
205.	Hexamethylbenzene	4.61	0	1	0
206.	Fluorine	4.18	0	1	0
207.	Benzohenone	3.18	0	1	0
208.	1,4,5-Trimethylnaphthalene	4.90	0	1	0
209.	Anthracene	4.45	0	1	0
210.	Phenanthrene	4.46	0	1	0
211.	1,2-Diphenylethane	4.79	0	1	0
212.	9-Methylanthracene	5.07	0	1	0
213.	Fluoranethene	5.16	0	1	0
214.	Pyrene	4.88	0	1	0
215.	9,10-Dimethylanthracene	5.69	0	1	0
216.	Diisobutylphthalate	4.11	0	1	0
217.	Benzo(*a*)fluorene	5.68	0	1	0
218.	Benzo(*b*)fluorene	5.77	0	1	0
219.	Benzo(*a*)anthracene	5.79	0	1	0
220.	Chirysene	5.73	0	1	0
221.	Triphenylene	5.49	0	1	0
222.	Benzo(*a*)pyrene	5.97	0	1	0
223.	Benzo(*ghi*)perylene	6.63	0	1	0

$Ip_1 = 1$ if halogens present in compounds, otherwise 0.
$Ip_2 = 1$ if benzene ring is present in compounds, otherwise 0.
$Ip_3 = 1$ if OH group is present in compounds, otherwise 0.

The added parameter Sz is the modification of W for cyclic compounds/cycle containing compounds. Its positive coefficient in Eq. 2 further indicates that the presence of aromatic ring is favorable for the exhibition of log *P*. The meaning of other terms is the same as discussed for Eq. 1.

The quality of Eq. 2 is further improved by the addition of Ip_1 to Eq. 2. The penta-parametric regression so yielded is found as below:

$$\begin{aligned}\log P = {} & 0.5165 + 0.2519(\pm 0.1041)^1\chi \\ & + 9.4891 \times 10^{-4}(\pm 5.3241 \times 10^{-4})\mathrm{Sz} \\ & + 0.7943(\pm 0.1804)\mathrm{Ip}_1 + 1.3909(\pm 0.2385)\mathrm{Ip}_2 \\ & - 1.2678(0.2256)\mathrm{Ip}_3 \\ n = 223, \quad & \mathrm{Se} = 1.1656, \quad R = 0.7979, \\ R^2_{\mathrm{A}} = 0.6282, \quad & F = 76.017, \quad Q = 0.6845 \end{aligned} \tag{3}$$

The coefficient of added parameter Ip_1 is positive in the above Eq. 3. This means that the presence of halogen is favorable for the exhibition of logs *P*. The contributions of other terms is the same as discussed above.

Step-wise regression has indicated that addition of χ_{eq} is still advantageous, thus improving the quality of regression as below:

Table 2. Calculated parameters for the series of aliphatic compounds used in the present study

Compd no	W	$^1\chi$	J	Sz	log RB	χ_{eq}
1	9	1.7321	2.3238	9	2.0794	2.9422
2	16	2.0000	3.0237	16	4.1589	3.1209
3	16	2.0000	3.0237	16	4.1589	2.9149
4	16	2.0000	3.0237	16	4.1589	2.7343
5	4	1.4142	1.6330	4	0.6931	2.4711
6	1	1.0000	1.0000	1	0.0000	2.5082
7	4	1.4142	1.6330	4	0.6931	2.5296
8	1	1.0000	1.0000	1	0.0000	2.3963
9	29	2.6427	2.9935	29	8.5533	2.6923
10	58	3.2500	4.0204	58	18.2053	3.5345
11	18	2.2701	2.5395	18	4.9698	2.5796
12	42	3.5412	2.9434	42	12.8300	2.6314
13	1	1.0000	1.0000	1	0.0000	2.3404
14	10	1.9142	1.9747	10	2.4849	2.4759
15	29	2.6427	2.9935	29	8.5533	2.6713
16	29	2.6427	2.9935	29	8.5533	2.5496
17	9	1.7321	2.3238	9	2.0794	2.7311
18	18	2.2701	2.5395	18	4.9698	2.5762
19	16	2.0000	3.0237	16	4.1589	2.4729
20	4	1.4142	1.6330	4	0.6931	2.4132
21	1	1.0000	1.0000	1	0.0000	2.2917
22	10	1.9142	1.9747	10	2.4849	2.3817
23	10	1.9142	1.9747	10	2.4849	2.4006
24	4	1.4142	1.6330	4	0.6931	2.3235
25	4	1.4142	1.6330	4	0.6931	2.3325
26	10	1.9142	1.9747	10	2.4849	2.4984
27	18	2.2701	2.5395	18	4.9698	2.4450
28	1	1.0000	1.0000	1	0.0000	2.2680
29	4	1.4142	1.6330	4	0.6931	2.3603
30	10	1.9142	1.9747	10	2.4849	2.4398
31	9	1.7321	2.3238	9	2.0794	2.3965
32	10	1.9142	1.9747	10	2.4849	2.3615
33	111	4.1547	4.2311	111	36.0572	3.0808
34	4	1.4142	1.6330	4	0.6931	2.8862
35	15	2.5000	2.0833	20	3.4657	2.3541
36	32	2.7701	2.6272	32	9.5342	2.4106
37	32	2.7701	2.6272	32	9.5342	2.4106
38	27	3.0000	2.0000	54	7.4547	2.4106
39	20	2.4142	2.1906	20	5.6630	2.3148
40	35	2.9142	2.3391	35	10.4505	2.3683
41	29	2.6427	2.9935	29	8.5533	2.3813
42	10	1.9142	1.9747	10	2.4849	2.2781
43	20	2.4142	2.1906	20	5.6630	2.3324
44	20	2.4142	2.1906	20	5.6630	2.3324
45	32	2.7701	2.6272	32	9.5342	2.3821
46	29	2.6427	2.9935	29	8.5533	3.2821
47	20	2.4142	2.1906	20	5.6630	2.3113
48	20	2.4142	2.1906	20	5.6630	2.3113
49	56	3.4142	2.4475	56	17.0297	2.3308
50	64	3.9319	2.1250	109	19.6969	2.5397
51	42	3.3939	2.1229	78	12.4245	2.4384
52	20	2.4142	2.1906	20	5.6630	2.3064
53	15	2.5000	2.0833	20	3.4657	2.2917
54	31	2.8081	2.7542	31	9.2465	2.1958
55	50	3.3081	2.8318	50	15.4203	2.3887
56	75	3.8081	2.9196	75	23.3858	2.4316
57	35	2.9142	2.3391	35	10.4505	2.3107
58	16	2.0000	3.0237	16	4.1589	2.2805
59	20	2.4142	2.1906	20	5.6630	2.2805
60	32	2.7701	2.6272	32	9.5342	2.3255
61	15	2.5000	2.0833	20	3.4657	2.5221
62	4	1.4142	1.6330	4	0.6931	2.3193
63	15	2.5000	2.0833	20	2.4657	2.4567
64	10	1.9142	1.9747	10	2.4849	2.4214
65	18	2.2701	2.5395	18	4.9698	2.5072
66	3	1.5000	2.2500	3	0.0000	2.2917
67	4	1.4142	1.6330	4	0.6931	2.2917
68	20	2.4142	2.1906	20	5.6630	2.3632
69	9	1.7321	2.3238	9	2.0794	2.3736
70	10	1.9142	1.9747	10	2.4849	2.3736
71	18	2.2701	2.5395	18	4.9698	2.4452
72	20	2.4142	2.1906	20	5.6630	2.3882
73	20	2.4142	2.1906	20	5.6630	2.4433
74	18	2.2701	2.5395	18	4.9698	2.4048
75	4	1.4142	1.6330	4	0.6931	2.2744
76	9	1.7321	2.3238	9	2.0794	2.3428
77	10	1.9142	1.9747	10	2.4849	2.3428
78	18	2.2701	2.5395	18	4.9698	2.4039
79	61	3.7877	2.2306	108	18.7806	2.5850
80	42	3.3939	2.1229	78	12.4245	2.5956
81	27	3.0000	2.0000	54	7.4547	2.4610
82	61	3.7877	2.2306	108	18.7806	2.5891
83	15	2.5000	2.0833	20	3.4657	2.4299
84	15	2.5000	2.0833	20	3.4657	2.3836
85	61	3.7877	2.2306	108	18.7806	2.5360
86	10	1.9142	1.9747	10	2.4849	2.3109
87	10	1.9142	1.9747	10	2.4849	2.3109
88	20	2.4142	2.1906	20	5.6630	2.3291
89	71	3.5607	3.1118	71	22.1049	2.5866
90	20	2.4142	2.1906	20	5.6630	2.3509
91	10	1.9142	1.9747	10	2.4849	2.2917
92	10	1.9142	1.9747	10	2.4849	2.2917
93	9	1.7321	2.3238	9	2.0794	2.2917
94	18	2.2701	2.5395	18	4.9698	2.3541
95	10	1.9142	1.9747	10	2.4849	2.3541
96	42	3.3939	2.1229	78	12.4245	2.3795
97	62	3.7879	2.1924	110	19.0038	2.4396
98	42	3.3939	2.1229	78	12.4245	2.3858
99	42	3.3939	2.1229	78	12.4245	2.4407
100	120	4.6984	2.2599	192	37.8252	2.5631
101	27	3.0000	2.0000	54	7.4547	2.3404
102	27	3.0000	2.0000	54	7.4547	2.3193
103	27	3.0000	2.0000	54	7.4547	2.3193
104	52	3.2701	2.6783	52	15.9311	2.3726
105	35	2.9141	2.3391	35	10.4505	2.3036
106	35	2.9141	2.3391	35	10.4505	2.3036
107	27	3.0000	2.0000	54	7.4547	2.3036
108	35	2.9142	2.3391	35	10.4505	2.2917
109	27	3.0000	2.0000	54	7.4547	2.2917
110	26	2.8939	2.1841	33	7.0493	2.2917
111	42	3.3939	2.1230	78	12.4245	2.3341
112	79	3.7701	2.7158	79	24.3021	2.3736
113	42	3.3939	2.1229	78	12.4245	2.3171
114	67	3.7187	3.3549	67	21.2655	2.3549
115	35	2.9142	2.3391	35	10.4505	2.2822
116	52	3.2701	2.6783	52	15.9311	2.3206
117	46	3.0607	3.1545	46	14.0985	2.3206
118	212	5.2701	2.7964	212	62.4563	2.3175
119	50	3.3081	2.8318	50	15.4203	2.3206
120	48	3.1259	2.9532	48	14.7917	2.3206
121	84	3.9142	2.5301	84	25.5549	2.3567
122	174	5.4641	2.7603	282	57.0114	2.6415
123	140	5.0366	2.6254	230	45.4623	2.5861
124	109	4.6259	2.5158	182	35.0118	2.5329
125	110	4.6091	2.4873	183	35.2995	2.5329
126	111	4.6091	2.4620	186	35.5227	2.5329
127	84	4.1984	2.3462	144	26.4585	2.4819
128	84	4.1815	2.3409	144	26.5230	2.4819
129	60	3.8045	2.2794	106	18.4930	2.4199

(*continued on next page*)

Table 2 (*continued*)

Compd no	W	$^1\chi$	J	Sz	log RB	χ_{eq}
130	62	3.7877	2.1924	110	19.0038	2.4199
131	61	3.7877	2.2306	108	18.7806	2.4328
132	60	3.8045	2.2794	106	18.4930	2.4328
133	62	3.7877	2.1924	110	19.0038	2.4328
134	114	4.7152	2.3960	180	36.5035	2.5765
135	88	4.3045	2.2284	142	27.6625	2.4279
136	88	4.3045	2.2284	142	27.6625	2.4279
137	42	3.3939	2.1229	78	12.4245	2.3305
138	64	3.9319	2.1250	109	19.6969	2.3802
139	64	3.9319	2.1250	109	19.6969	2.3802
140	60	3.8045	2.2790	106	18.4930	2.3802
141	42	3.5000	2.0417	63	12.5423	2.2917
142	59	3.7071	2.3279	104	18.0875	2.3282
143	56	3.4142	2.4475	56	17.0297	2.2834
144	84	3.9142	2.5301	84	25.5549	2.3042
145	79	4.4663	1.9698	144	24.4436	2.4143
146	88	4.3045	2.2284	142	27.6625	2.3869
147	64	3.9319	2.1250	109	19.6969	2.3239
148	61	3.7877	2.2306	108	18.7806	2.3239
149	60	3.8045	2.2794	106	18.493	2.3239
150	62	3.7877	2.1924	110	19.0038	2.3239
151	258	6.5366	2.2264	423	84.4145	2.4981
152	117	4.8805	2.3400	184	37.4198	2.4047
153	88	4.3045	2.2284	142	27.6625	2.3467
154	84	3.9142	2.5301	84	25.5549	2.2843
155	108	4.1259	2.9147	108	33.3663	2.2945
156	220	5.4142	2.6909	220	64.0657	2.3695
157	84	4.1984	2.3462	144	26.4585	2.3193
158	94	4.4319	2.0779	148	29.2710	2.3326
159	164	5.3425	2.2050	236	51.6487	2.4091
160	82	4.2152	2.4133	140	25.9477	2.3193
161	84	4.1815	2.3409	144	26.5230	2.3193
162	88	4.3045	2.2284	142	27.6625	2.3193
163	94	4.4319	2.0779	148	29.2710	2.3193
164	120	4.6984	2.2599	192	37.8252	2.3394
165	490	7.7540	1.6972	664	144.8246	2.5537
166	109	4.9663	1.9254	243	34.4240	2.3571
167	327	6.6851	2.3058	474	100.1843	2.4772
168	215	5.7364	2.2008	323	66.4610	2.3938
169	252	5.9097	2.3924	378	78.7726	2.4067
170	111	4.6091	2.4620	186	35.5227	2.3158
171	125	4.8637	2.1738	200	39.0780	2.3158
172	114	4.6052	2.3892	177	36.3212	2.3158
173	133	4.9319	2.0172	196	41.1983	2.3158
174	75	3.8081	2.9196	75	23.3858	2.3567
175	79	3.7701	2.7158	79	24.3021	2.3567
176	120	4.4142	2.5951	120	36.1595	2.3907
177	52	3.2701	2.6783	52	15.9311	2.3059
178	56	3.4142	2.4475	56	17.0297	2.3059
179	48	3.3461	2.9923	48	14.9094	2.3059
180	354	6.9136	2.6408	516	113.0512	2.7230
181	64	3.9319	2.1250	109	19.6969	2.4095
182	79	4.4663	1.9698	144	24.4436	2.4301
183	64	3.9319	2.1250	109	19.6969	2.4084
184	88	4.3045	2.2284	142	27.6625	2.4595
185	64	3.9319	2.1250	109	19.6969	2.3663
186	88	4.3045	2.2284	142	27.6625	2.4279
187	140	5.0366	2.6254	230	45.4623	2.3131
188	414	7.5922	2.0139	705	130.7054	2.4565
189	412	7.6091	2.0247	704	130.1946	2.4565
190	352	7.1815	1.9604	606	111.1012	2.4301
191	354	7.1815	1.9473	612	111.5067	2.4301
192	348	7.1815	1.9869	600	110.3185	2.4301
193	293	6.7709	1.9208	512	92.8878	2.4042
194	286	6.7877	1.9728	498	91.4119	2.4042

Table 2 (*continued*)

Compd no	W	$^1\chi$	J	Sz	log RB	χ_{eq}
195	240	6.3770	1.8848	426	76.1739	2.3789
196	166	5.9495	1.8447	369	53.4924	2.3540
197	198	5.9663	1.7997	360	62.3223	2.3540
198	436	7.4495	1.5281	652	127.195	2.3881
199	179	5.7709	2.0150	372	57.6275	2.3404
200	176	5.7877	2.0549	366	56.9344	2.3404
201	176	5.7877	2.0526	366	56.9344	2.3404
202	182	5.7709	1.9762	378	58.2561	2.3404
203	186	5.7540	1.9355	386	59.1034	2.3404
204	446	7.7019	2.5508	588	138.9793	2.4377
205	174	5.4641	2.7603	282	57.0114	2.3169
206	219	6.4495	1.7616	420	70.0929	2.3601
207	307	6.8770	1.8180	505	95.8633	2.3925
208	216	6.1984	2.1226	438	70.3806	2.3349
209	279	6.9327	1.6825	656	88.4649	2.3656
210	271	6.9495	1.7402	632	86.8067	2.3656
211	94	4.4319	2.0779	148	29.2719	2.3404
212	334	7.3433	1.7136	766	106.0542	2.3565
213	364	7.9495	1.6770	801	117.8215	2.3757
214	362	7.9327	1.6706	1008	117.2825	2.3754
215	385	7.7709	1.7942	866	123.3716	2.3504
216	878	9.4136	2.5143	1080	256.1823	2.3941
217	453	8.4327	1.5895	940	143.7743	2.3665
218	453	8.4327	1.5895	940	143.7743	2.3665
219	553	8.9158	1.5122	1325	171.8145	2.3701
220	545	8.9327	1.5384	1301	170.1563	2.3701
221	513	8.9495	1.6420	1269	164.6027	2.3701
222	680	9.9158	1.4872	1887	214.1680	2.3784
223	815	10.9158	1.4749	2571	261.2488	2.3857

$$\begin{aligned}\log P &= 6.2290 + 0.2893(\pm 0.0998)^1\chi \\ &\quad + 8.8888 \times 10^{-4}(\pm 5.0894 \times 10^{-4})\mathrm{Sz} \\ &\quad - 2.4400(\pm 0.5246)\chi_{eq} + 1.1964(\pm 0.1929)\mathrm{Ip}_1 \\ &\quad + 1.2216(\pm 0.2308)\mathrm{Ip}_2 - 1.2091(\pm 0.2160)\mathrm{Ip}_3 \\ n &= 223, \quad \mathrm{Se} = 1.1139, \quad R = 0.8183, \\ R^2_{\mathrm{A}} &= 0.6605, F = 72.977, \quad Q = 0.7346\end{aligned} \tag{4}$$

Now, the added parameter χ_{eq} has negative coefficient in Eq. 4 indicating thereby that electronegativity is not favorable for the exhibition of log P. The physical significance of other terms is the same as discussed above.

Further regression analysis has indicated that by adding J to Eq. 4 yields the following expression with slightly improved quality:

$$\begin{aligned}\log P &= 6.5511 + 0.1720(\pm 0.1109)^1\chi \\ &\quad + 0.0016(\pm 5.8314 \times 10^{-4})\mathrm{Sz} \\ &\quad - 2.9180(\pm 0.5583)\chi_{eq} + 0.4929(\pm 0.2115)J \\ &\quad + 1.1709(\pm 0.1912)\mathrm{Ip}_1 + 1.3977(0.2406)\mathrm{Ip}_2 \\ &\quad - 1.2592(\pm 0.2149)\mathrm{Ip}_3 \\ n &= 223, \quad \mathrm{Se} = 1.1026, \quad R = 0.8233, \\ R^2_{\mathrm{A}} &= 0.6673, F = 64.611, \quad Q = 0.7466\end{aligned} \tag{5}$$

Table 3. Correlation matrix for the correlation of the molecular descriptors used log *P* of a series of compounds

	log *P*	*W*	$^1\chi$	*J*	Sz	log RB	Ip_1	Ip_2	Ip_3	χ_{eq}
log *P*	1.0000									
W	0.5957	1.0000								
$^1\chi$	0.6770	0.9304	1.0000							
J	0.1571	0.2244	0.1375	1.0000						
Sz	0.5972	0.9429	0.8773	0.3129	1.0000					
log RB	0.6065	0.9991	0.9371	0.2272	0.9507	1.0000				
Ip_1	0.1737	0.0746	0.0715	0.1845	0.0852	0.0719	1.0000			
Ip_2	0.6842	0.5606	0.7090	0.2256	0.5333	0.5708	0.0540	1.0000		
Ip_3	0.3862	0.1572	0.1728	0.1665	0.1636	0.1596	−0.0168	0.2141	1.0000	
χ_{eq}	0.1525	0.0240	0.0487	0.3995	0.0459	0.0236	0.4301	0.0936	0.0671	1.0000

Table 4. Regression parameters and quality of the proposed models

Compd no	Parameters used	A_i $i = 1,2,3,4,5,6,7,8$	*B*	R^2	*R*	Se	*F*-Ratio	$Q = R/Se$
1	$^1\chi$	0.3672(±0.0594)	0.4467	0.5990	0.7740	1.2187	109.053	0.6351
	Ip_2	1.4262(±0.2380)						
	Ip_3	−1.3032(±0.2354)						
2	$^1\chi$	0.2159(±0.1081)	0.8088	0.6041	0.7772	1.2138	83.162	0.6403
	Sz	$9.2761 \times 10^{-4}(\pm 5.5439 \times 10^{-4})$						
	Ip_2	1.5365(±0.2460)						
	Ip_3	−1.2785(±0.2349)						
3	$^1\chi$	0.2519(±0.1041)	0.5165	0.6366	0.7979	1.1656	76.017	0.6845
	Sz	$9.4891 \times 10^{-4}(\pm 5.3241 \times 10^{-4})$						
	Ip_1	0.7943(0.1804)						
	Ip_2	1.3909(±0.2385)						
	Ip_3	−1.2678(±0.2256)						
4	$^1\chi$	0.2893(±0.998)	6.2290	0.6697	0.8183	1.1139	72.977	0.7346
	Sz	$8.8888 \times 10^{-4}(\pm 5.0894 \times 10^{-4})$						
	χ_{eq}	−2.4400(±0.5246)						
	Ip_1	1.1964(±0.5246)						
	Ip_2	1.2216(±0.2308)						
	Ip_3	−1.2091(±0.2160)						
5	$^1\chi$	0.1720(±0.1109)	6.5511	0.6778	0.8233	1.1026	64.611	0.7466
	Sz	$0.0016(\pm 5.8314 \times 10^{-4})$						
	χ_{eq}	−2.9180(±0.5583)						
	J	0.4929(±0.2115)						
	Ip_1	1.1709(±0.1912)						
	Ip_2	1.3977(±0.2406)						
	Ip_3	−1.2592(±0.2149)						
6	$^1\chi$	0.2930(±0.1448)	6.2228	0.6803	0.8248	1.1009	56.924	0.8240
	W	−0.0028(±0.0021)						
	J	0.4648(±0.2123)						
	Sz	$0.0022(\pm 7.4200 \times 10^{-4})$						
	χ_{eq}	−2.8596(±0.5592)						
	Ip_1	1.1809(±0.1911)						
	Ip_2	1.2996(±0.2519)						
	Ip_3	−1.2573(±0.2145)						
7	$^1\chi$	0.1968(±0.0982)	5.1713	0.7519	0.8652	0.9226	87.452	0.9191
	Sz	$0.0015(5.0463 \times 10^{-4})$						
	J	0.4363(±0.1826)						
	χ_{eq}	−2.2749(±0.4772)						
	Ip_1	1.0666(±0.1623)						
	Ip_2	1.3080(±0.2128)						
	Ip_3	−1.3433(±0.1819)						
8	$^1\chi$	0.3172(±0.1292)	4.8429	0.7544	0.8686	0.9202	77.171	0.9439
	W	−0.0027(0.0019)						
	J	0.3989(±0.1840)						
	Sz	$0.0020(\pm 6.3368 \times 10^{-4})$						
	χ_{eq}	−2.2060(±0.4784)						
	Ip_1	1.0781(±0.1621)						
	Ip_2	1.1993(±0.2255)						
	Ip_3	−1.3446(±0.1815)						

Table 5. Comparison of estimated lipophilicity ($\log P$) with their observed values

Compd no	Obsd. $\log P$	Model-8	
		Estimated $\log P$	Residue
1	1.080	0.901	0.179
2	2.160	0.867	1.293
3	2.530	1.321	1.209
4	2.830	1.720	1.110
5	1.250	1.567	−0.317
6	0.350	0.025	0.325
7	−1.510	0.360	−1.870
8	−0.770	−1.072	0.302
9	3.400	1.996	1.404
10	2.000	0.723	1.277
11	2.610	1.952	0.658
12	3.220	2.388	0.832
13	0.370	0.396	−0.026
14	2.090	1.848	0.242
15	0.920	2.043	−1.123
16	2.390	2.311	0.079
17	1.240	1.367	−0.127
18	0.410	1.960	−1.550
19	2.490	2.296	0.194
20	−3.400	0.617	−4.017
21	1.130	0.503	0.627
22	1.960	2.056	−0.096
23	1.480	2.014	−0.534
24	1.610	1.893	−0.283
25	1.430	1.873	−0.443
26	−0.760	0.454	−1.214
27	1.810	0.555	1.255
28	−0.310	−0.611	0.301
29	−1.360	−0.495	−0.865
30	−1.310	−0.317	−0.993
31	1.660	0.717	0.943
32	0.250	−0.426	0.676
33	0.460	1.275	−0.815
34	0.730	1.432	−0.702
35	0.790	1.432	−0.642
36	−0.270	1.313	−1.583
37	2.640	2.442	0.198
38	0.850	1.188	−0.338
39	2.760	1.206	1.554
40	0.880	−0.020	0.900
41	0.890	1.325	−0.435
42	−0.830	0.150	−0.980
43	−0.920	−1.727	0.807
44	0.580	1.372	−0.792
45	0.730	1.372	−0.642
46	−1.430	0.381	−1.811
47	−0.310	1.388	−1.698
48	1.420	2.513	−1.093
49	2.480	1.382	1.098
50	3.000	1.412	1.588
51	0.990	1.969	−0.979
52	1.240	1.722	−0.482
53	1.210	1.805	−0.595
54	2.520	2.659	−0.139
55	3.110	1.643	1.467
56	3.390	1.439	1.951
57	1.190	0.275	0.915
58	0.120	0.904	−0.784
59	0.940	0.824	0.116
60	−0.080	1.048	−1.128
61	−0.380	0.890	−1.270
62	0.350	1.034	−0.684
63	1.720	1.159	0.561
64	1.770	0.885	0.885
65	2.180	2.335	−0.155
66	−0.240	1.078	−1.318
67	0.170	−0.349	0.519
68	0.330	1.171	−0.841
69	0.500	0.935	−0.435
70	−0.280	0.814	−1.094
71	−1.010	1.260	−2.270
72	2.360	0.923	1.437
73	0.050	−0.199	0.249
74	0.250	−0.281	0.531
75	−0.920	−0.083	−0.837
76	1.530	2.368	−0.838
77	−0.230	1.202	−1.432
78	1.340	1.108	0.232
79	1.990	1.134	0.856
80	1.460	1.134	0.326
81	2.530	2.410	0.120
82	0.710	1.197	−0.487
83	0.530	1.284	−0.754
84	2.400	1.176	1.224
85	2.330	1.176	1.154
86	2.340	1.258	1.082
87	0.290	1.372	−1.082
88	0.880	1.039	−0.159
89	2.990	3.842	−0.852
90	2.590	2.530	0.060
91	2.890	3.828	−0.938
92	2.270	3.707	−1.437
93	1.910	1.508	0.402
94	2.130	2.667	−0.537
95	2.470	1.514	0.956
96	2.300	1.514	0.786
97	0.580	1.683	−1.103
98	2.870	1.597	1.273
99	2.730	1.597	1.133
100	2.860	1.549	1.311
101	3.390	1.623	1.767
102	3.440	1.575	1.865
103	3.370	1.575	1.795
104	1.230	0.320	0.910
105	1.780	1.837	−0.057
106	2.470	1.702	0.768
107	0.340	2.124	−1.784
108	1.760	0.453	1.307
109	1.480	0.580	0.900
110	1.650	2.387	−0.737
111	1.650	0.527	1.123
112	1.520	1.864	−0.344
113	0.830	0.498	0.332
114	5.730	4.240	1.490
115	5.180	4.157	1.023
116	4.640	4.085	0.555
117	4.660	4.068	0.592
118	4.600	4.061	0.539
119	4.020	3.983	0.037
120	4.190	3.976	0.214
121	3.640	3.955	−0.315
122	3.790	3.918	−0.128
123	3.530	3.903	−0.373
124	3.430	3.926	−0.496
125	3.440	3.889	−0.449
126	1.910	2.608	−0.698
127	0.340	1.797	−1.457
128	1.310	1.797	−0.487
129	2.730	2.872	−0.142

Table 5 (*continued*)

Compd no	Obsd. log *P*	Model-8	
		Estimated log *P*	Residue
130	2.110	2.939	−0.829
131	1.100	1.594	−0.494
132	1.960	1.620	0.340
133	1.330	0.522	0.808
134	2.570	1.959	0.611
135	2.670	3.003	−0.333
136	1.580	3.087	−1.507
137	3.150	3.063	0.087
138	3.200	3.065	0.135
139	3.120	3.089	0.031
140	3.150	3.051	0.099
141	1.600	3.283	−1.683
142	2.310	3.176	−0.866
143	2.780	2.186	0.594
144	0.560	0.926	−0.366
145	3.780	3.264	0.516
146	3.230	3.183	0.047
147	2.640	3.348	−0.708
148	3.660	3.293	0.367
149	3.420	3.257	0.163
150	3.660	3.236	0.424
151	3.720	3.213	0.507
152	2.850	3.346	−0.496
153	1.340	2.398	−1.058
154	3.300	3.392	−0.092
155	2.250	3.716	−1.466
156	3.420	3.547	−0.127
157	3.890	3.663	0.227
158	4.000	3.462	0.538
159	4.450	3.419	1.031
160	4.110	3.406	0.704
161	4.380	3.349	1.031
162	0.840	1.970	−1.130
163	0.750	0.530	0.220
164	−0.670	0.586	−1.256
165	1.760	1.830	−0.070
166	1.670	1.781	−0.111
167	1.450	1.982	−0.532
168	1.550	3.394	−1.844
169	1.560	2.874	−1.314
170	1.320	2.968	−1.648
171	1.480	2.877	−1.397
172	1.870	2.927	−1.057
173	2.300	4.047	−1.747
174	0.850	1.797	−0.947
175	4.560	3.682	0.878
176	5.780	5.251	0.529
177	6.410	5.264	1.146
178	5.760	5.120	0.640
179	5.900	5.122	0.778
180	5.710	5.129	0.581
181	5.160	4.997	0.163
182	4.980	5.013	−0.033
183	4.530	4.878	−0.348
184	3.920	3.784	0.136
185	4.010	3.668	0.342
186	4.210	3.918	0.292
187	4.420	3.797	0.623
188	4.370	3.814	0.556
189	4.380	3.813	0.567
190	4.400	3.786	0.614
191	4.310	3.770	0.540
192	4.610	3.878	0.732
193	4.180	3.859	0.321
194	3.180	3.885	−0.705

Table 5 (*continued*)

Compd no	Obsd. log *P*	Model-8	
		Estimated log *P*	Residue
195	4.900	4.024	0.876
196	4.450	4.291	0.159
197	4.460	4.292	0.168
198	4.790	3.166	1.624
199	5.070	4.535	0.535
200	5.160	4.659	0.501
201	4.880	5.080	−0.200
202	5.690	4.782	0.908
203	4.110	4.619	−0.509
204	5.680	4.845	0.835
205	5.770	4.845	0.925
206	5.790	5.479	0.311
207	5.730	5.467	0.263
208	5.490	5.534	−0.044
209	5.970	6.578	−0.608
210	6.630	7.911	−1.281

Finally, when all the parameters are used in the regression analysis still better quality model, as given below, is obtained:

$$\begin{aligned}\log P &= 6.2228 + 0.2930(\pm 0.1448)^{1}\chi \\ &\quad - 0.0028(\pm 0.0021)W + 0.4648(\pm 0.2123)J \\ &\quad + 0.0022(\pm 7.4200 \times 10^{-4})\mathrm{Sz} \\ &\quad - 2.8596(\pm 0.5592)\chi_{\mathrm{eq}} + 1.1809(\pm 0.1911)\mathrm{Ip}_1 \\ &\quad + 1.2996(\pm 0.2519)\mathrm{Ip}_2 - 1.2579(\pm 0.2145)\mathrm{Ip}_3 \\ n &= 223, \quad \mathrm{Se} = 1.0009, \quad R = 0.8248, \\ R^2_{\mathrm{A}} &= 0.6684, F = 56.924, \quad Q = 0.8240 \end{aligned} \tag{6}$$

From the aforementioned results we conclude that Eqs. 5 and 6 is best suitable for modeling log *P*. We have, therefore, estimated log *P* values using both these equations and found that compounds **27**, **32**, **41**, **80**, **82**, **84**, **85**, **115**, **141**, **143**, **151**, **154**, and **204** (= **13**) gave quite a large residues (difference between observed and calculated log *P*) and are, therefore, considered as outliners. This means that either these compounds have different types of mechanism or they are behaving differently in regression procedure adopted by us. Whatever be the reason the deletion of these 13 compounds represents only 5% (approximate) calling of the data. The deletion of these 13 compounds from the regression procedure results into two regression expressions with better *R*-value: (i) $^{1}\chi$, *W*, *J*, χ_{eq}, Ip_1, Ip_2, Ip_3 and (ii) $^{1}\chi$, Sz, *J*, χ_{eq}, Ip_1, Ip_2, Ip_3. The former regression expression was not considered, as the coefficient of *W* term was quite smaller than its standard deviation. Such expressions are not allowed statistically. The latter expression was found as below:

$$\begin{aligned}\log P &= 5.1713 + 0.1968(\pm 0.0982)^{1}\chi \\ &\quad + 0.0015(\pm 5.0463 \times 10^{-4})\mathrm{Sz} \\ &\quad + 0.4363(\pm 0.1826)J - 2.2749(\pm 0.4772)\chi_{\mathrm{eq}} \\ &\quad + 1.0666(\pm 0.1623)\mathrm{Ip}_1 + 1.3080(\pm 0.2128)\mathrm{Ip}_2 \\ &\quad - 1.3433(\pm 0.1819)\mathrm{Ip}_3 \\ n &= 210, \quad \mathrm{Se} = 0.9413, \quad R = 0.8652, \\ R^2_{\mathrm{A}} &= 0.7328, F = 82.866, \quad Q = 0.9191 \end{aligned} \tag{7}$$

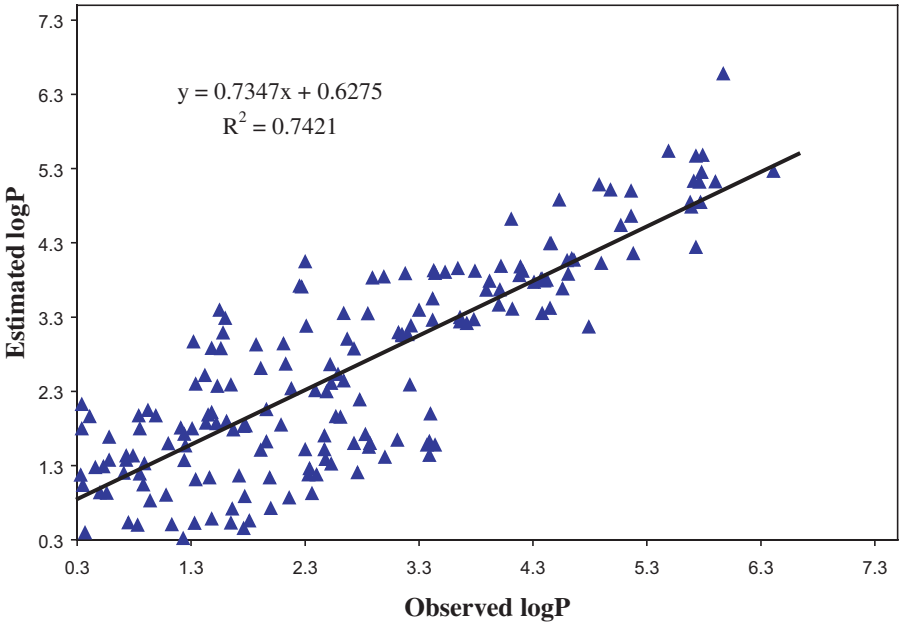

Figure 1. Comparison of observed and estimated lipophilicity (log *P*) using model-8.

Finally, when all the eight parameters are used in the regression, we obtain still better model:

$$\begin{aligned}\log P &= 4.8429 + 0.3172(\pm 0.5292)^{1}\chi \\ &\quad - 0.0027(\pm 0.0019)W + 0.3989(\pm 0.1840)J \\ &\quad + 0.0020(\pm 6.3368 \times 10^{-4})\mathrm{Sz} \\ &\quad - 2.2060(\pm 0.4784)\chi_{eq} + 1.0781(\pm 0.1625)\mathrm{Ip}_1 \\ &\quad + 1.1993(\pm 0.2255)\mathrm{Ip}_2 - 1.3446(\pm 0.1815)\mathrm{Ip}_3 \\ n &= 210, \quad \mathrm{Se} = 0.9202, \quad R = 0.8686, \\ R_A^2 &= 0.7496, \quad F = 77.171, \quad Q = 0.9439 \end{aligned} \tag{8}$$

In order to confirm our findings we have estimated log *P* using Eq. 8 and compared them with the observed value of log *P*. Such a comparison is shown in the Table 5. We then obtained predictive correlation coefficient of the model from the correlation of observed and estimated log *P* (Fig. 1), which comes out to be (R^2_{pred}) 0.7421. This clearly indicates that this model is the best model for estimating log *P* of present set of compounds.

We have also made attempts to investigate predictive power of the proposed models by using quality factor (*Q*). This is done by using Pogliani's method,[24,25] which defines *Q* as the ratio of correlation coefficient (*R*) to the standard error of estimation (Se), that is, $Q = R/\mathrm{Se}$. Thus, the larger the value of *Q* the better will be the predictive power of the model. We observed that calculated *Q* values goes on increasing as we pass from Eq. 1–8.

Further supports in favor of our results are obtained by estimating other interesting statistical parameters viz. probable error of the coefficient of correlation (PE), index of forecasting (*E*), and coefficient of alience (*k*). The calculations of these parameters are described in our earlier publications[26–28] and some such calculations are also available in the literature.[23] The calculated values of these parameters along with the values of R^2 and R^2_A are given in Table 6.

It is argued that if $R < \mathrm{PE}$ then the correlation is not significant. The value of $R > \mathrm{PE}$, at least three times greater, indicates that there exists appreciable correlation. Finally, if $R > 6\mathrm{PE}$ then it indicates that the correlation is definitely good. The perusal of Table 6 shows that all the proposed correlation are good and that the last correlation (Eq. 8) is the best for modeling lipophilicity. Table 6 also shows the index of forecasting efficiency (*E*), as expected, increases with correlation coefficient. The higher value of correlation coefficient ($R > 0.5$) and its relation to the coefficient of alienation (*k*) indicates that the degree of relationship is greater than the degree of lack of relationship.

Table 6. Calculated values of PE, *E*, and *k* for the proposed models

Model (eqn)	R^2	R^2_A	6PE	*E*	*k*
1.	0.5991	0.5935	0.1611	36.68	0.6332
2.	0.6040	0.5968	0.1061	37.08	0.6293
3.	0.6366	0.6282	0.0974	39.72	0.6028
4.	0.6696	0.6605	0.0886	42.52	0.5758
5.	0.6778	0.6673	0.0863	43.24	0.5676
6.	0.6803	0.6684	0.0856	43.46	0.5654
7.	0.7487	0.7328	0.0694	49.86	0.5014
8.	0.7896	0.7496	0.0587	54.13	0.4587

It is also important to comment on R^2_A. It is a measure of the % explained variation in the dependent variables that takes into account the relationship between the number of cases and the number of independent variables in the regression model. Whereas R^2 will always increase when an independent variable is added. R^2_A on the other hand will decrease if the added variable does not reduce the unexplained variation enough to offset the loss of degrees of freedom. Therefore, if a variable is added that does not contribute its fair share, the R^2_A will actually decline. Table 6 shows that there is consistent increase in R^2_A as we pass from three variable to eight variable models. We attempted regression analysis for the sub-group generated by the collection of the outliner's mention above. We failed to obtain a good model. However, the results obtained herein provides additional support to our earlier work[29–32] related to the modeling of lipophilicity on a border set now containing a very large set of heterogeneous set of compounds.

3. Conclusions

The results have shown that the combination of some of the distance-based topological indices with equalized electronegativity and indicator parameters is an alternative method for modeling monitoring and estimating lipophilicity/hydrophobicity ($\log P$) of a large heterogeneous set of compounds used by us. However, still there is need to obtain a single topological index, which is capable of modeling lipophilicity. Attempts in this regard is underway.

4. Experimental

4.1. Lipophilicity

The lipophilicity ($\log P$) of the heterogeneous set of 223 compounds is chosen from the compilation made by Hansch.[1]

4.2. Topological indices

All the distance-based topological indices are calculated from the hydrogen suppressed molecular graphs obtained by deleting all the carbon–hydrogen as well as heteroatom–hydrogen bonds from the respective molecular structures.[2] These indices are calculated using the software provided by Lukovits.

4.3. Regression analysis

The stepwise regression analysis[23] was carried out using maximum R^2.

Acknowledgements

Authors are thankful to Professor Istvan Lukovits, Hungarian Academy of Sciences, Budapest, Hungary for providing software to carryout regression analysis and to Professor Ivan Gutman, Faculty of Science, University of Kragujevac, Yugoslavia for introducing one of the authors (P.V.K.) to this fascinating field of Chemical Graph Theory and Topology. We are also thankful to the Council of Scientific and Industrial Research (CSIR), New Delhi, India for the financial support.

References and notes

1. Agrawal, V. K.; Singh, J.; Khadikar, P. V. *Bioorg. Med. Chem.* **2002**, *10*, 3981.
2. Agrawal, V. K.; Bano, S.; Khadikar, P. V. *Bioorg. Med. Chem.* **2003**, *11*, 4039.
3. Agrawal, V. K.; Sohgaura, R.; Khadikar, P. V. *Bioorg. Med. Chem.* **2001**, *9*, 3295.
4. Agrawal, V. K.; Khadikar, P. V. *Bioorg. Med. Chem.* **2002**, *10*, 3571.
5. Agrawal, V. K.; Chaturvedi, S.; Khadikar, P. V.; Abraham, M. H. *Bioorg. Med. Chem.* **2003**, *11*, 4523.
6. Agrawal, V. K.; Karmarkar, S.; Khadikar, P. V.; Shrivastva, S.; Lukovits, I. *Indian J. Chem.* **2003**, *42A*, 1426.
7. Khadikar, P. V.; Srivastava, A.; Agrawal, V. K.; Shrivastava, S. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 3009.
8. Agrawal, V. K.; Khadikar, P. V. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 447.
9. Balaban, A. T.; Basak, S. C.; Beteringhe, A.; Mills, D.; Supuran, C. T. *Mol. Divers.* **2004**, *8*, 401.
10. Agrawal, V. K.; Mishra, K.; Khadikar, P. V. *Oxid. Commun.* **2003**, *26*, 14.
11. Wiener, H. *J. Am. Chem. Soc.* **1947**, *69*, 17.
12. Khadikar, P. V.; Deshpandey, N. V.; Kale, P. P.; Dubrynin, A.; Gutman, I.; Domotor, G. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 547.
13. Gutman, I.; Khadikar, P. V.; Rajput, P. V.; Karmarkar, S. *J. Serb. Chem. Soc.* **1997**, *62*, 235.
14. Randic, M. *Graph Theoretical Approach to Structure–Activity Studies: Search for Optimal Antitumor Compounds in Molecular Basis of Cancer Part A, Macromolecular Structure, Carcinogens and Oncogens*; Alan R, NC: New York, 1985; pp 309–318.
15. Kier, L. B.; Hall, L. H. *Molecular Connectivity in Structure Activity Relationship*; Wiley: New York, 1986; Kier, L. B.; Hall, L. H. *Molecular Structure Description*; Academic: New York, 1999.
16. Kier, L. B. *J. Pharm. Sci.* **1980**, *69*, 807–810.
17. Gombar, V. K.; Enslein, K. *Quant. Struct.–Act. Relat.* **1997**, *16*, 309.
18. Kier, L. B.; Hall, L. H. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 1039.
19. Khadikar, P. V.; Karmarkar, S.; Agrawal, V. K. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 934.
20. Khadikar, P. V.; Agrawal, V. K.; Karmarkar, S. *Oxid. Commun.* **2004**, *27*, 17.
21. Todeschini, R.; Consonni, V. *Handbook of Molecular Descriptors*; Wiley-VCH: Weinheim (GER), 2000; Karelson, M. *Molecular Descriptors in QSAR/QSPR*; Wiley: New York, 2000; *QSPR/QSAR Studies by Molecular Descriptors*; Diudea, M. V., Ed.; Babes-Bolyai University, Cluj: Romania, 2000; Trinajstic, N. *Chemical Graph Theory*, 2nd ed.; CRC: Boca Raton, FL, 1992.
22. Khadikar, P. V.; Lukovits, I.; Agrawal, V. K.; Shrivastva, S.; Jaiswal, M.; Gutman, I.; Karmarkar, S.; Shrivastava, A. *Indian J. Chem.* **2003**, *42A*, 1436.
23. Chaterjee, S.; Hadi, A. S.; Price, B. *Regression Analysis by Examples*, 3rd ed.; Wiley: New York, 2000.

24. Pogliani, L. J. *Phys. Chem.* **1994**, *98*, 1494.
25. Pogliani, L. *Chem. Rev.* **2000**, *100*, 3827.
26. Khadikar, P. V.; Sharma, V.; Karmarkar, S.; Supuran, C. T. *Bioorg. Med. Chem. Lett*. **2005**, *15*, in press.
27. Khadikar, P. V.; Sharma, V.; Verma, R. G. *Bioorg. Med. Chem. Lett*. **2005**, *15*, in press.
28. Jaiswal, M.; Khadikar, P. V.; Sisodia, R. S. *Bioorg. Med. Chem. Lett*. **2005**, *15*, in press.
29. Jaiswal, M.; Khadikar, P. V.; Supuran, C. T. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 5661.
30. Thakur, M.; Agrawal, A.; Thakur, A.; Khadikar, P. V. *Bioorg. Med. Chem.* **2004**, *12*, 2287.
31. Khadikar, P. V.; Singh, S.; Shrivastava, A. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 1125.
32. Khadikar, P. V.; Agrawal, V. K.; Karmarkar, S. *Bioorg. Med. Chem.* **2002**, *10*, 3499.